

# Use of Markov Chain Monte Carlo to estimate uncertainties in Bayesian reconstructions based on deformable models

**Kenneth M. Hanson**

*Los Alamos National Laboratory*

Presentation available under <http://home.lanl.gov/kmh/>

# Acknowledgements

---

- MCMC experts
  - ▶ Julian Besag, Jim Guberantus, John Skilling, Malvin Kalos, David Higdon
- Deformable models
  - ▶ Greg Cunningham, Xavier Battle
- General discussions
  - ▶ Greg Cunningham, Richard Silver

# Overview

---

- Overview of Markov Chain Monte Carlo (MCMC) technique
  - ▶ for drawing random samples from probability density functions
- Bayesian approach to model-based analysis
- Example - tomographic reconstruction from two views
  - ▶ Deformable geometric models
- Probabilistic interpretation of priors (MCMC)
- Estimation of uncertainty in reconstructed shape
  - ▶ Use of MCMC to sample posterior
  - ▶ Hard truth approach - probe model stiffness

# MCMC - problem statement

---

- Parameter space of  $n$  dimensions represented by vector  $\mathbf{x}$
- Given an “arbitrary” **target** probability density function (pdf),  $q(\mathbf{x})$ , draw a set of samples  $\{\mathbf{x}_k\}$  from it
- Only requirement typically is that, given  $\mathbf{x}$ , one be able to evaluate  $Cq(\mathbf{x})$ , where  $C$  is an unknown constant
  - ▶ MCMC algorithms do not typically require knowledge of the normalization constant of the target pdf; from now on the multiplicative constant  $C$  will not be made explicit
- Although focus here is on continuous variables, MCMC applies to discrete variables as well
- Called a Markov chain since  $\mathbf{x}_{k+1}$  depends only on  $\mathbf{x}_k$

# Uses of MCMC

---

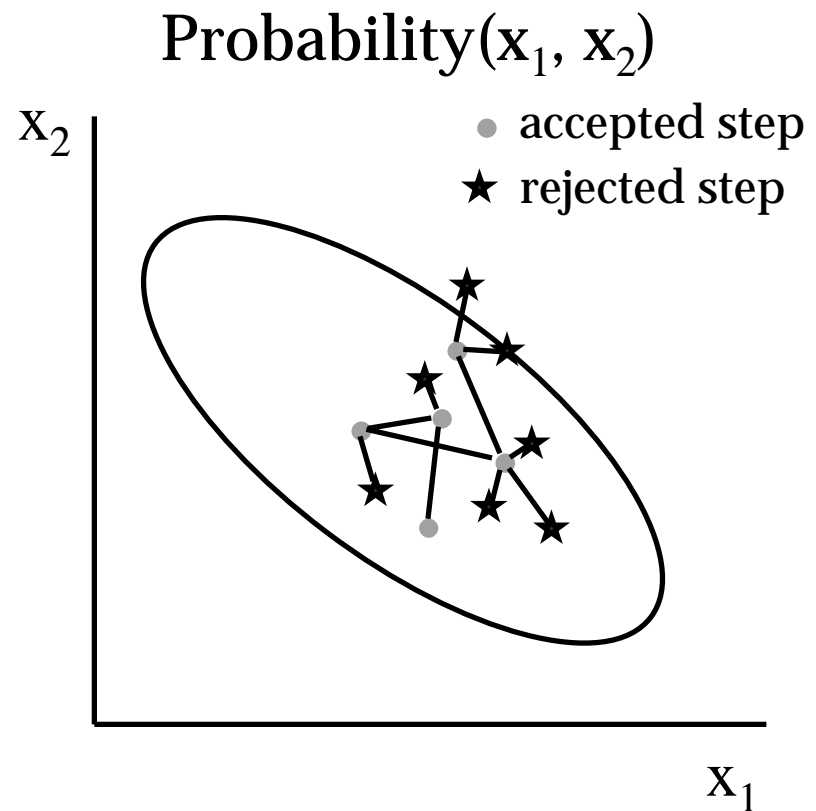
- Permits evaluation of the expectation values
  - ▶ for K samples,  $\langle f(\mathbf{x}) \rangle = \int f(\mathbf{x}) q(\mathbf{x}) d\mathbf{x} \cong (1/K) \sum_k f(\mathbf{x}_k)$
  - ▶ typical use is to calculate mean  $\langle \mathbf{x} \rangle$  and variance  $\langle (\mathbf{x} - \langle \mathbf{x} \rangle)^2 \rangle$
- Useful for evaluating integrals, such as the partition function for properly normalizing the pdf
- Dynamic display of sequence as video loop
  - ▶ provides visualization of uncertainties in model and range of model variations
- Automatic marginalization
  - ▶ when considering any subset of parameters of an MCMC sequence, the remaining parameters are marginalized over

# Markov Chain Monte Carlo

---

Generates sequence of random samples from an arbitrary probability density function

- Metropolis algorithm:
  - ▶ draw trial step from symmetric pdf, i.e.,  
 $t(\Delta \mathbf{x}) = t(-\Delta \mathbf{x})$
  - ▶ accept or reject trial step
  - ▶ simple and generally applicable
  - ▶ relies only on calculation of target pdf for any  $\mathbf{x}$



# Metropolis algorithm

---

- Select initial parameter vector  $\mathbf{x}_0$
- Iterate as follows: at iteration number  $k$ 
  - (1) create new trial position  $\mathbf{x}^* = \mathbf{x}_k + \Delta\mathbf{x}$  ,  
where  $\Delta\mathbf{x}$  is randomly chosen from  $t(\Delta\mathbf{x})$
  - (2) calculate ratio  $r = q(\mathbf{x}^*)/q(\mathbf{x}_k)$
  - (3) accept trial position, i.e. set  $\mathbf{x}_{k+1} = \mathbf{x}^*$   
if  $r \geq 1$  or with probability  $r$ , if  $r < 1$   
otherwise stay put,  $\mathbf{x}_{k+1} = \mathbf{x}_k$
- Requires only computation of  $q(\mathbf{x})$

# Choice of trial distribution

---

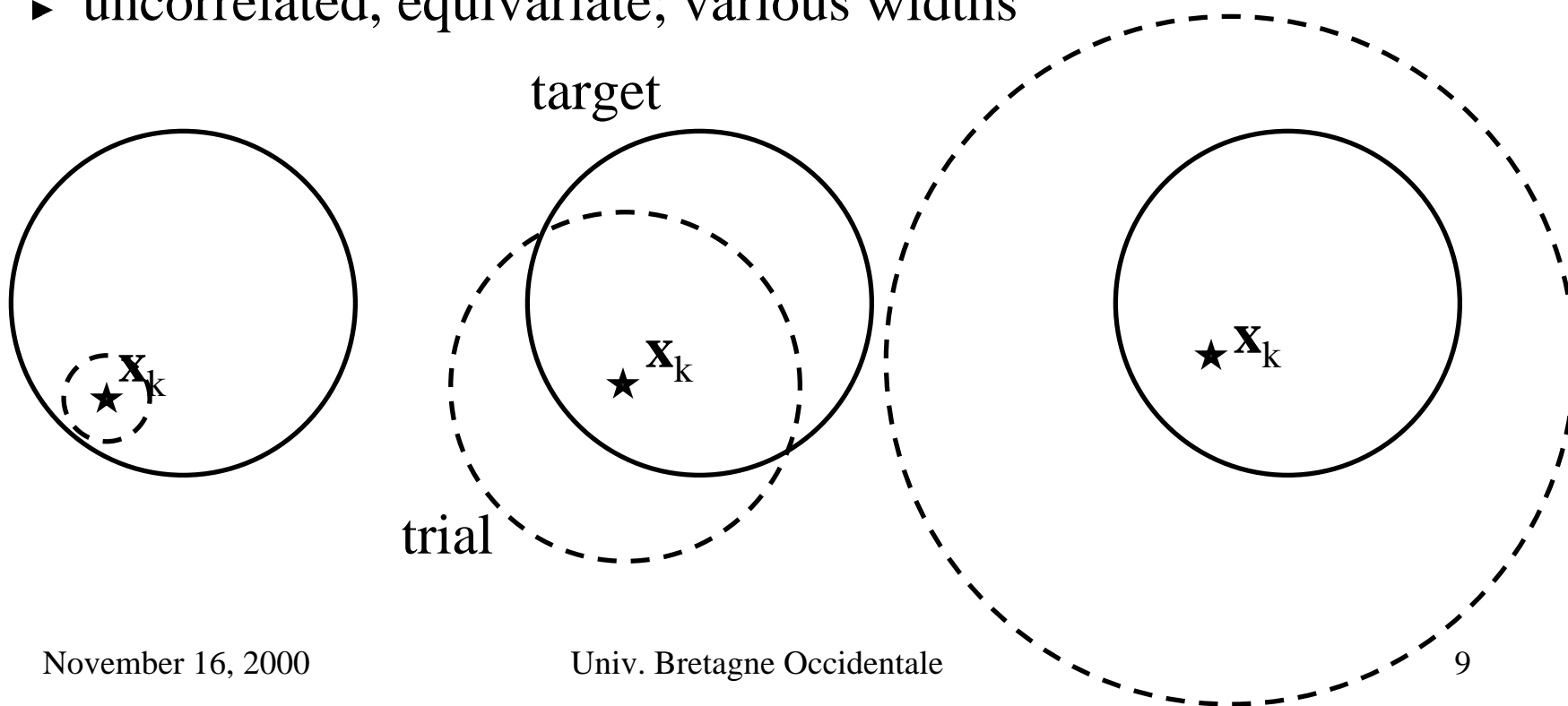
- Loose requirements on trial distribution  $t(\cdot)$ 
  - ▶ stationary; independent of position
- Often used functions include
  - ▶  $n$ -D Gaussian, isotropic and uncorrelated
  - ▶  $n$ -D Cauchy, isotropic and uncorrelated
- Choose width to “optimize” MCMC efficiency
  - ▶ rule of thumb: aim for acceptance fraction of about 25%



# Experiments with the Metropolis algorithm

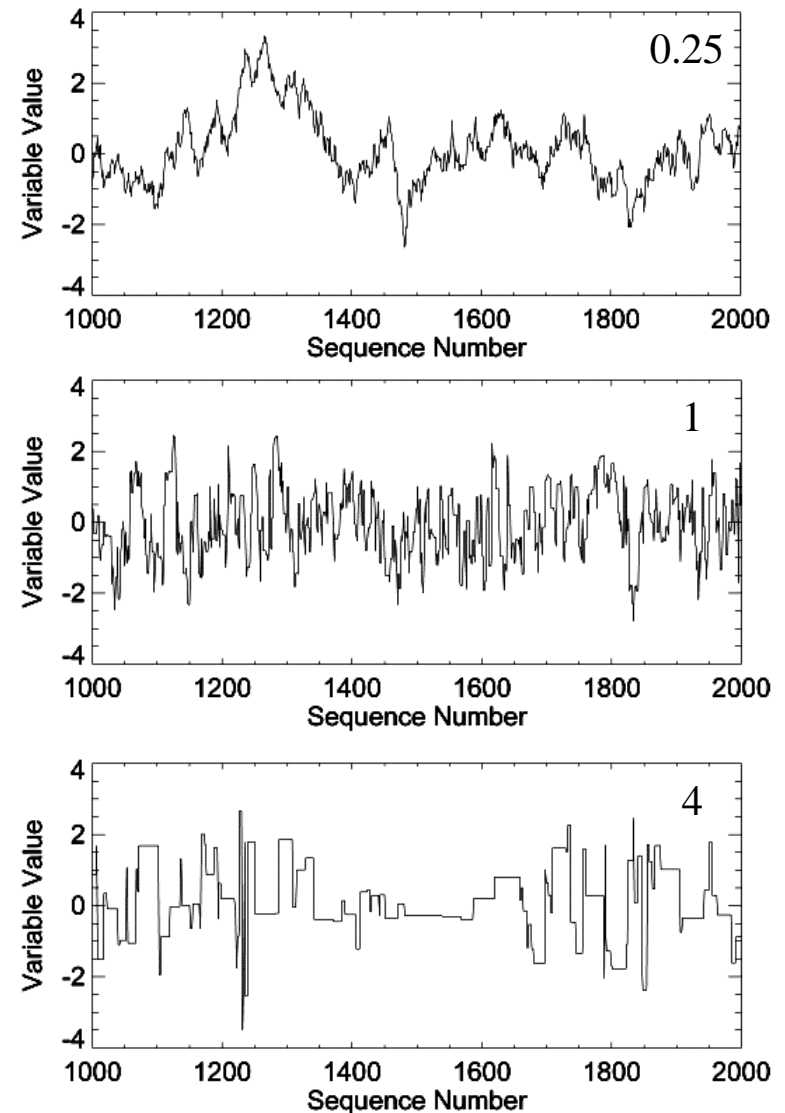
---

- Target distribution  $q(\mathbf{x})$  is  $n$  dimensional Gaussian
  - ▶ uncorrelated, univariate (isotropic with unit variance)
  - ▶ most generic case
- Trial distribution  $t(\Delta\mathbf{x})$  is  $n$  dimensional Gaussian
  - ▶ uncorrelated, equivariate; various widths



# MCMC sequences for 2D Gaussian

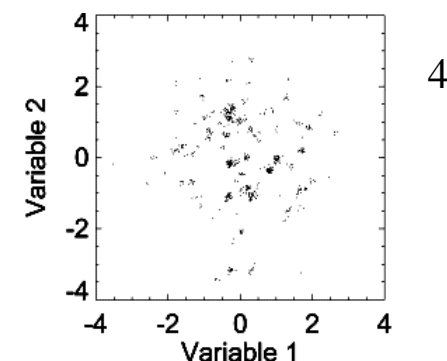
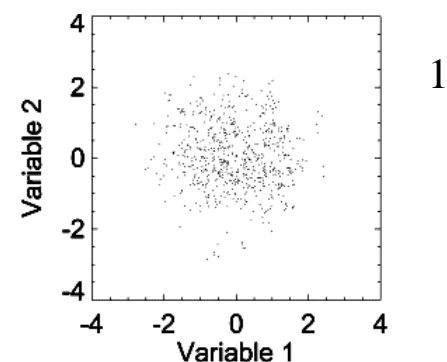
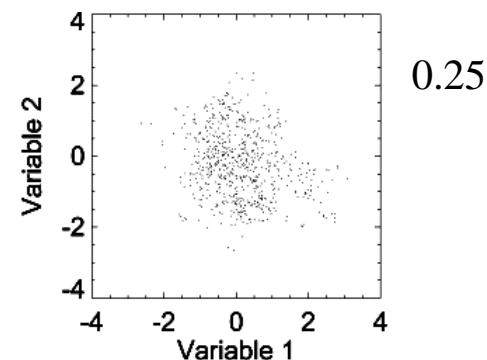
- ▶ results of running Metropolis with ratios of width of trial to target of 0.25, 1, and 4
- ▶ when trial pdf is much smaller than target pdf, movement across target pdf is slow
- ▶ when trial width same as target, samples seem to sample target pdf better
- ▶ when trial width much larger than target, trials stay put for long periods, but jumps are large
  - this example from Hanson and Cunningham (SPIE, 1998)



# MCMC sequences for 2D Gaussian

---

- ▶ results of running Metropolis with ratios of width of trial to target of 0.25, 1, and 4
- ▶ display accumulated 2D distribution for 1000 trials
- ▶ viewed this way, it is difficult to see difference between top two images
- ▶ when trial pdf much larger than target, fewer splats, but further apart



# MCMC - autocorrelation and efficiency

---

- ▶ In MCMC sequence, subsequent parameter values are usually correlated

- ▶ Degree of correlation quantified by **autocorrelation function**:

$$\rho(l) = \frac{1}{N} \sum_{i=1}^N y(i) y(i-l)$$

where  $y(x)$  is the sequence and  $l$  is lag

- ▶ For Markov chain, expect exponential

$$\rho(l) = \exp\left[-\left|\frac{l}{\lambda}\right|\right]$$

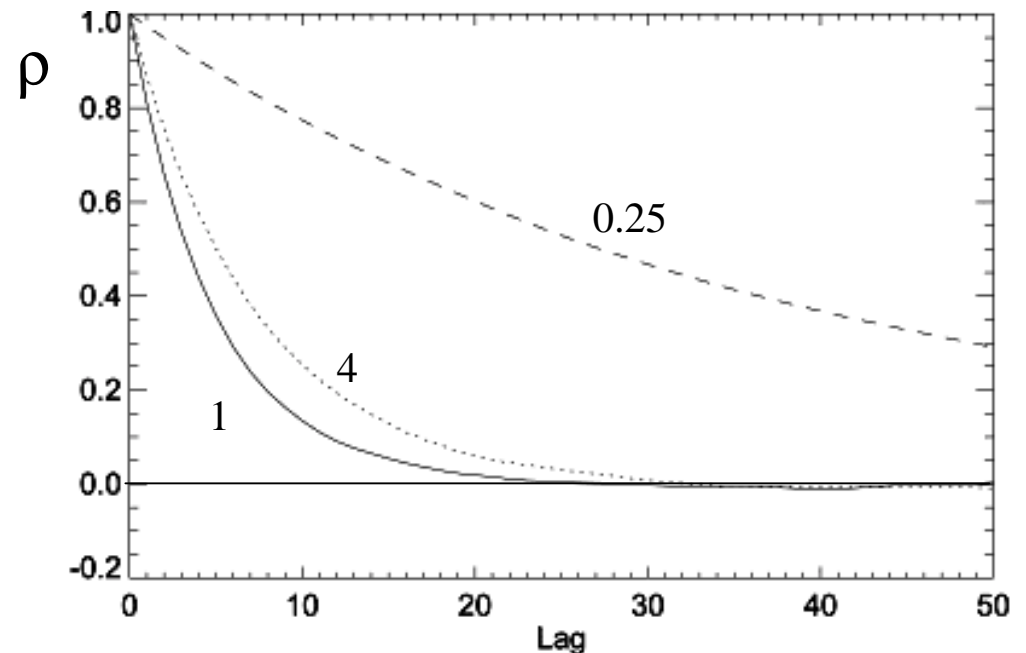
- ▶ Sampling **efficiency** is

$$\eta = \left[1 + 2 \sum_{l=1}^{\infty} \rho(l)\right]^{-1} = \frac{1}{1 + 2\lambda}$$

- ▶ In other words,  $\eta^{-1}$  iterates required to achieve one statistically independent sample

# Autocorrelation for 2D Gaussian

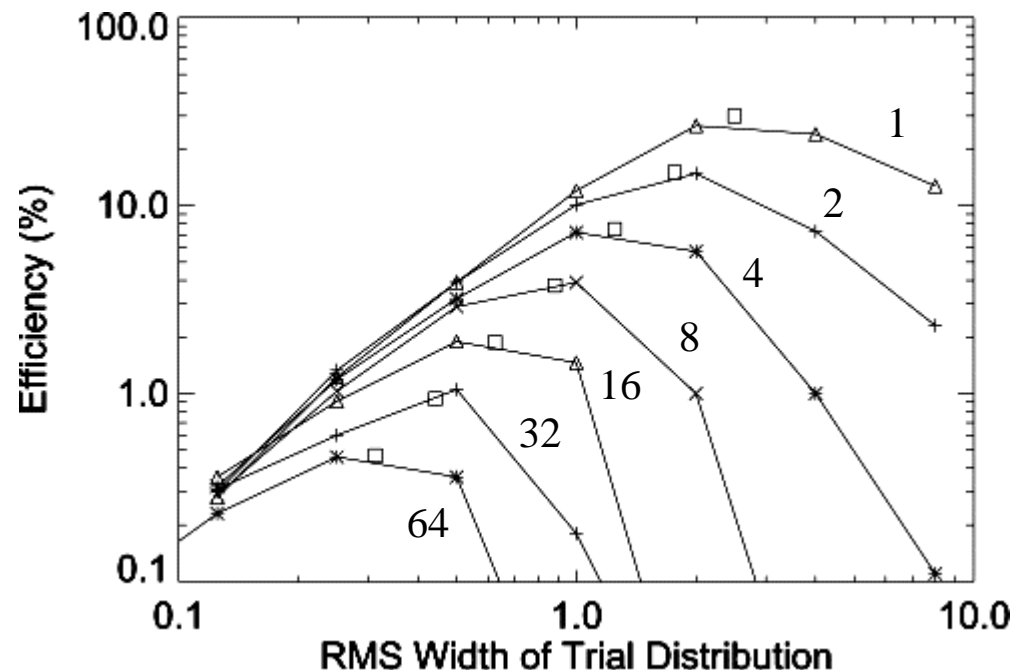
- ▶ plot confirms that the autocorrelation drops slowly when the trial width is much smaller than the target width; MCMC efficiency is poor
- ▶ best efficiency is when trial width about same as target width (for 2D)



Normalized autocovariance for  
various widths of trial pdf  
relative to target: 0.25, 1, and 4

# Efficiency as function of width of trial pdf

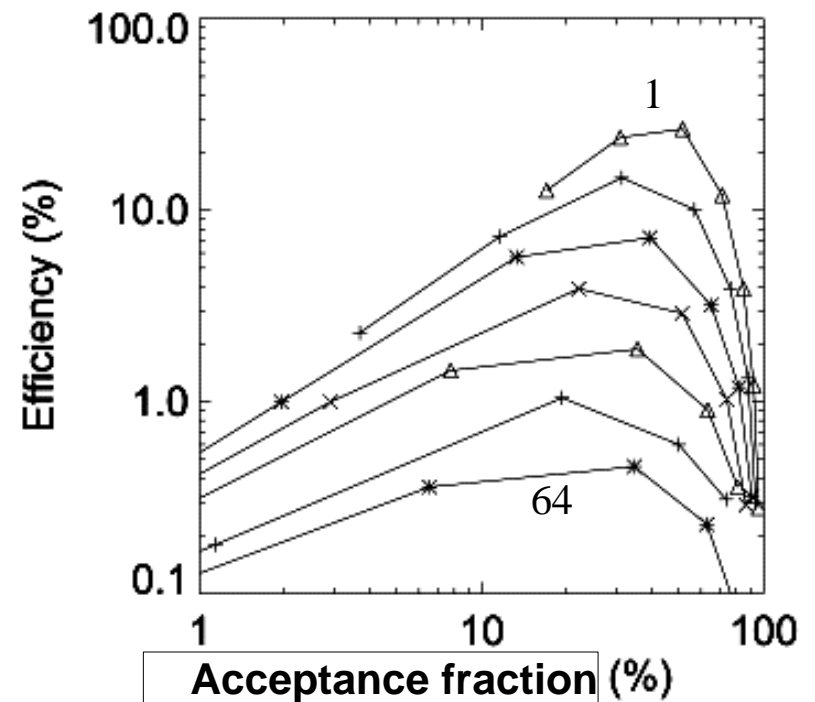
- ▶ for univariate Gaussians, with 1 to 64 dimensions
- ▶ efficiency as function of width of trial distributions
- ▶ boxes are predictions of optimal efficiency from diffusion theory [A. Gelman, et al., 1996]
- ▶ efficiency drops reciprocally with number of dimensions



# Efficiency as function of acceptance fraction

---

- ▶ for univariate Gaussians, with 1 to 64 dimensions
- ▶ efficiency as function of acceptance fraction
- ▶ best efficiency is achieved when about 25% of trials are accepted for a moderate number of dimensions



# Efficiency of Metropolis algorithm

---

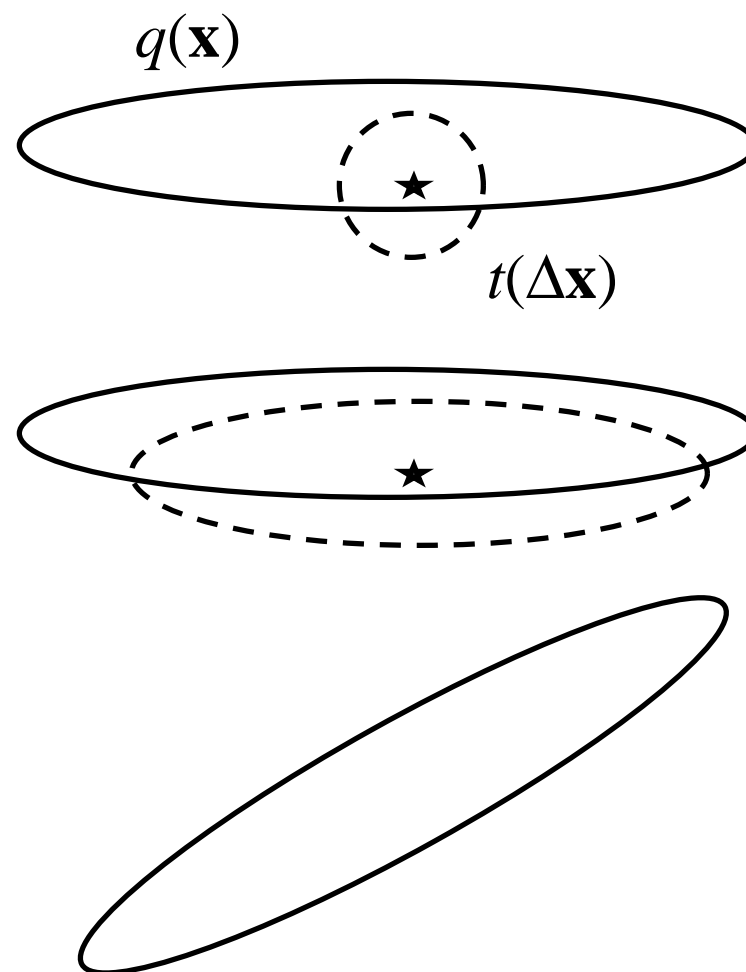
- Results of experimental study agree with predictions from diffusion theory (A. Gelman et al., 1996)
- Optimum choice for width of Gaussian trial distribution occurs for acceptance fraction of about 25% (but is a weak function of number of dimensions)
- Optimal statistical efficiency:  $\eta \sim 0.3/n$ 
  - ▶ holds for simplest case of uncorrelated, equivariate Gaussian
  - ▶ correlation and variable variance generally decreases efficiency



# Further considerations

---

- When target distribution  $q(\mathbf{x})$  not isotropic
  - ▶ difficult to accommodate with isotropic  $t(\Delta\mathbf{x})$
  - ▶ each parameter can have different efficiency
  - ▶ desirable to vary width of different  $t(\mathbf{x})$  to approximately match  $q(\mathbf{x})$
  - ▶ recovers efficiency of univariate case
- When  $q(\mathbf{x})$  has correlations
  - ▶  $t(\mathbf{x})$  should match shape of  $q(\mathbf{x})$



# MCMC Issues

---

- Confirmation of **convergence** to target pdf
  - ▶ is sequence in thermodynamic equilibrium with target pdf?
  - ▶ validity of estimated properties of parameters (covariance)
- **Burn in**
  - ▶ at beginning of sequence, may need to run MCMC for awhile to achieve convergence to target pdf
- Use of **multiple sequences**
  - ▶ different starting values can help confirm convergence
  - ▶ natural choice when using computers with multiple CPUs
- **Accuracy** of estimated properties of parameters
  - ▶ related to efficiency, described above
- Optimization of **efficiency** of MCMC

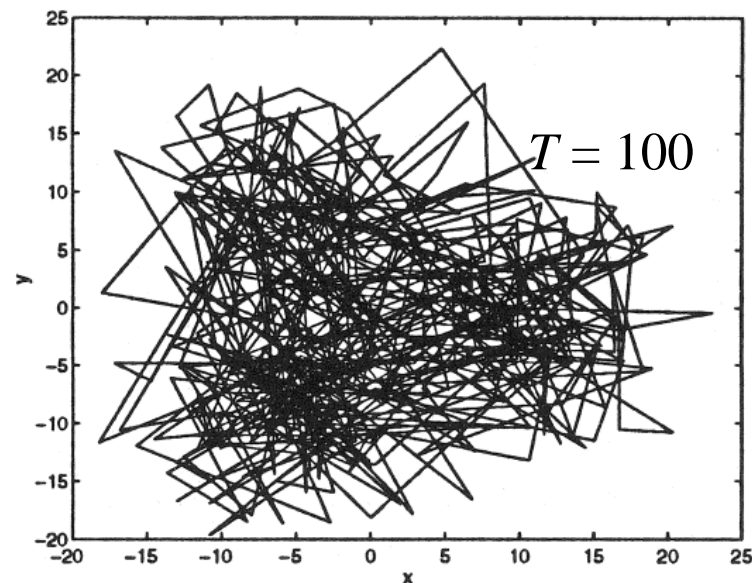
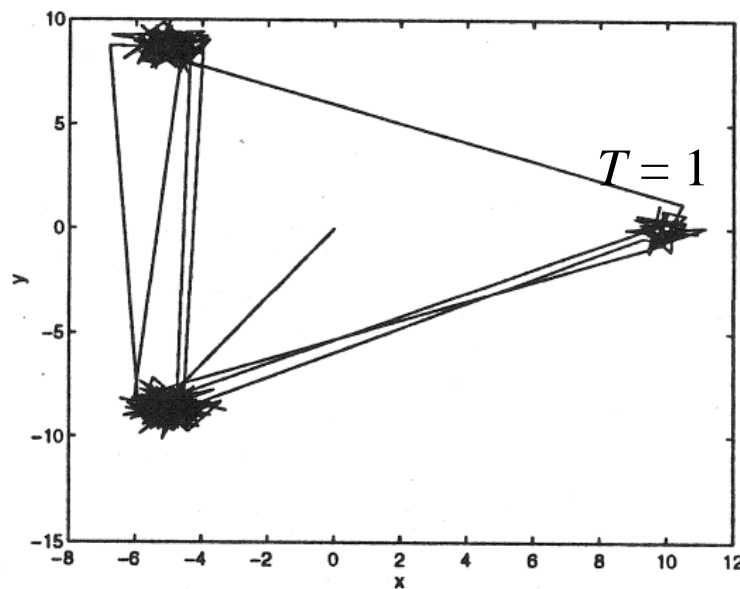
# Annealing

---

- Introduction of fictitious **temperature**
  - ▶ define functional  $\varphi(\mathbf{x})$  as minus-logarithm of target probability
$$\varphi(\mathbf{x}) = -\log(q(\mathbf{x}))$$
  - ▶ scale  $\varphi$  by an inverse “temperature” to form new pdf
$$q^\dagger(\mathbf{x}, T) = \exp[-T^{-1} \varphi(\mathbf{x})]$$
  - ▶  $q^\dagger(\mathbf{x}, T)$  is flatter than  $q(\mathbf{x})$  for  $T > 1$  (called annealing)
- Uses of annealing (also called tempering)
  - ▶ allows MCMC to move between multiple peaks in  $q(\mathbf{x})$
  - ▶ simulated annealing optimization algorithm (takes  $\lim T \rightarrow 0$ ) for purpose of finding global minimum
  - ▶ estimate normalization constant (**partition function**) by including  $T$  as parameter in MCMC:
$$Z = \int q(\mathbf{x}) d\mathbf{x} = \int_0^1 \int q^\dagger(\mathbf{x}, T) d\mathbf{x} dT$$

# Annealing to handle multiple peaks

- ▶ Example - target distribution is three narrow, well-separated peaks
- ▶ For original distribution ( $T = 1$ ), an MCMC run of 10000 steps rarely moves between peaks
- ▶ At temperature  $T = 100$  (right), MCMC moves easily between peaks and through surrounding regions



from M-D Wu and W. J. Fitzgerald, *Maximum Entropy and Bayesian Methods* (1996)

# Other MCMC algorithms

---

- **Gibbs**

- ▶ vary only one component of  $\mathbf{x}$  at a time
- ▶ draw new value of  $x_j$  from conditional  $q(x_j | x_1 x_2 \dots x_{j-1} x_{j+1} \dots )$

- **Metropolis-Hastings**

- ▶ allows use of nonsymmetric trial functions,  $t(\Delta\mathbf{x}; \mathbf{x}_k)$ , suitably chosen to improve efficiency
- ▶ use  $r = [t(\Delta\mathbf{x}; \mathbf{x}_k) q(\mathbf{x}^*)] / [t(-\Delta\mathbf{x}; \mathbf{x}^*) q(\mathbf{x}_k)]$

- **Langevin technique**

- ▶ uses gradient\* of minus-log-prob to shift trial function towards regions of higher probability
- ▶ uses Metropolis-Hastings

\* **adjoint differentiation** provides efficient gradient calculation

# Hamiltonian hybrid algorithm

---

- ▶ called hybrid because it alternates Gibbs & Metropolis steps
- ▶ associate with each parameter  $x_i$  a **momentum**  $p_i$

- ▶ define a Hamiltonian

$$H = \varphi(\mathbf{x}) + \sum p_i^2 / (2 m_i) \quad ; \quad \text{where } \varphi = -\log (q(\mathbf{x}))$$

- ▶ new pdf:

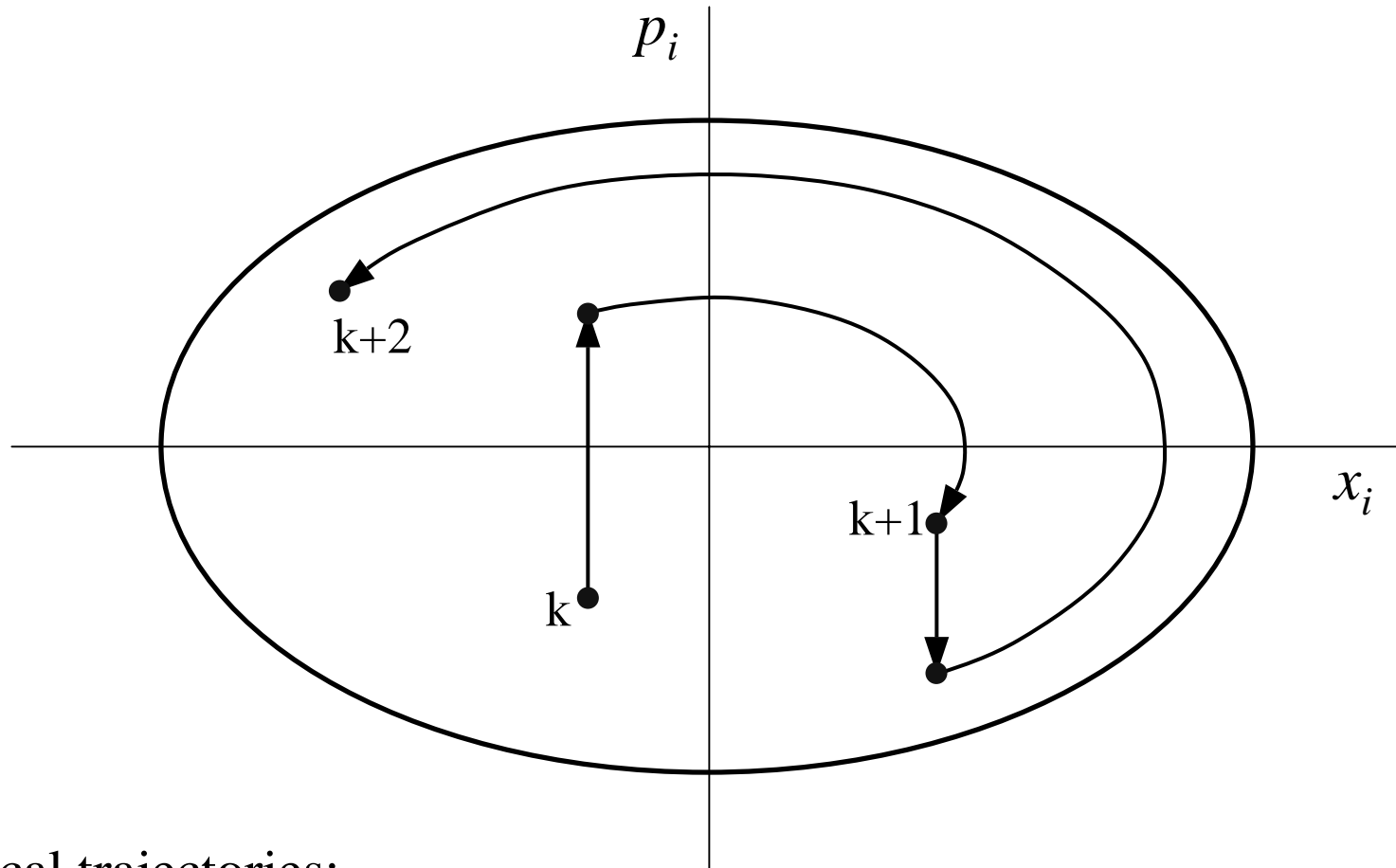
$$q'(\mathbf{x}, \mathbf{p}) = \exp(-H(\mathbf{x}, \mathbf{p})) = q(\mathbf{x}) \exp(-\sum p_i^2 / (2 m_i))$$

- ▶ can easily move long distances in  $(\mathbf{x}, \mathbf{p})$  space at **constant  $H$**  using **Hamiltonian dynamics**, so Metropolis step is very efficient
  - requires gradient\* of  $\varphi$  (minus-log-prob)
- ▶ Gibbs step: draw  $\mathbf{p}$  from Gaussian (at fixed  $\mathbf{x}$ )
- ▶ efficiency may be better than Metropolis for large dimensions

\* **adjoint differentiation** provides efficient gradient calculation

# Hamiltonian hybrid algorithm

---



Typical trajectories:

red path - Gibbs sample from momentum distribution

green path - trajectory with constant  $H$ , followed by Metropolis

# Conclusions about MCMC

---

- MCMC provides good tool for exploring the posterior and hence for drawing inferences about models and parameters
- For valid results, care must be taken to
  - ▶ verify convergence of the sequence
  - ▶ exclude early part of sequence, before convergence reached
  - ▶ be wary of multiple peaks that need to be sampled
- For good efficiency, care must be taken to
  - ▶ adjust the size and shape of the trial distribution; rule of thumb is to aim for 25% trial acceptance for  $5 < n < 100$
- A lot of research is happening - don't worry, be patient



# Bayesian approach to model-based analysis

---

- **Models**

- ▶ used to describe and analyze physical world
- ▶ parameters inferred from data

- **Bayesian analysis**

- ▶ uncertainties in parameters described by probability density functions (pdf)
- ▶ prior knowledge about situation may be incorporated
- ▶ quantitatively and logically consistent methodology for making inferences about models
- ▶ open-ended approach
  - can incorporate new data
  - can extend models and choose between alternatives

# Bayesian viewpoint

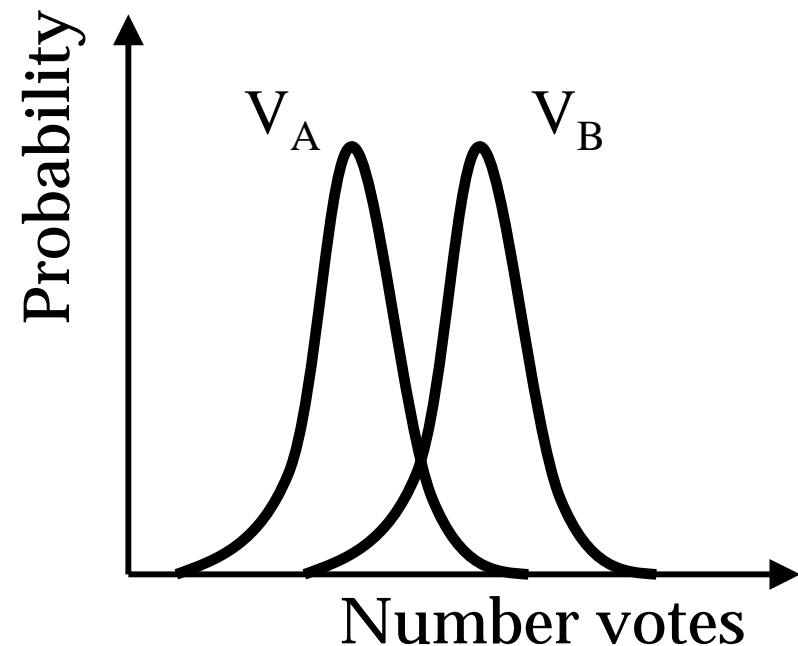
---

- Focus on probability distribution functions (pdf)
  - ▶ uncertainties in estimates more important than the estimates themselves
- Bayes law:  $p(\mathbf{a}|\mathbf{d}) \sim p(\mathbf{a}) p(\mathbf{d}|\mathbf{a})$ 
  - ▶ where  $\mathbf{a}$  is parameter vector and  $\mathbf{d}$  represents data
  - ▶ pdf before experiment,  $p(\mathbf{a})$  (called *prior*)
  - ▶ modified by pdf describing experiments,  $p(\mathbf{d}|\mathbf{a})$  (*likelihood*)
  - ▶ yields pdf summarizing what is known,  $p(\mathbf{a}|\mathbf{d})$  (*posterior*)
- Experiment should provide decisive information
  - ▶ posterior much narrower than prior

# Who wins the election?

---

- Process: people vote for candidate A or candidate B
  - ▶  $V_A$  = number of votes A receives
  - ▶  $V_B$  = number of votes B receives
- Winner is one with simple majority
  - ▶ if  $V_A > V_B$ , A wins, etc.
- Before election, pollsters sample voters; try to predict who will win
- Plot shows B ahead of A: but considering uncertainties, “it is too close to call”



# Who wins the election?

---

- During voting process, one can combine known results with predictions for unknown results to obtain new prediction for outcome
  - ▶ should arrive at more narrow probability distributions
- After voting process, one knows  $V_A$  and  $V_B$  with certainty: winner declared!

# Bayesian model building

---

- Steps in model building
  - ▶ choose how to model (represent) object
  - ▶ assign priors to parameters based on what is known beforehand
  - ▶ for given measurements, determine model with highest posterior probability (MAP)
  - ▶ assess uncertainties in model parameters
- Higher levels of inference
  - ▶ assess suitability of model to explain data
  - ▶ if necessary, try alternative models and decide among them

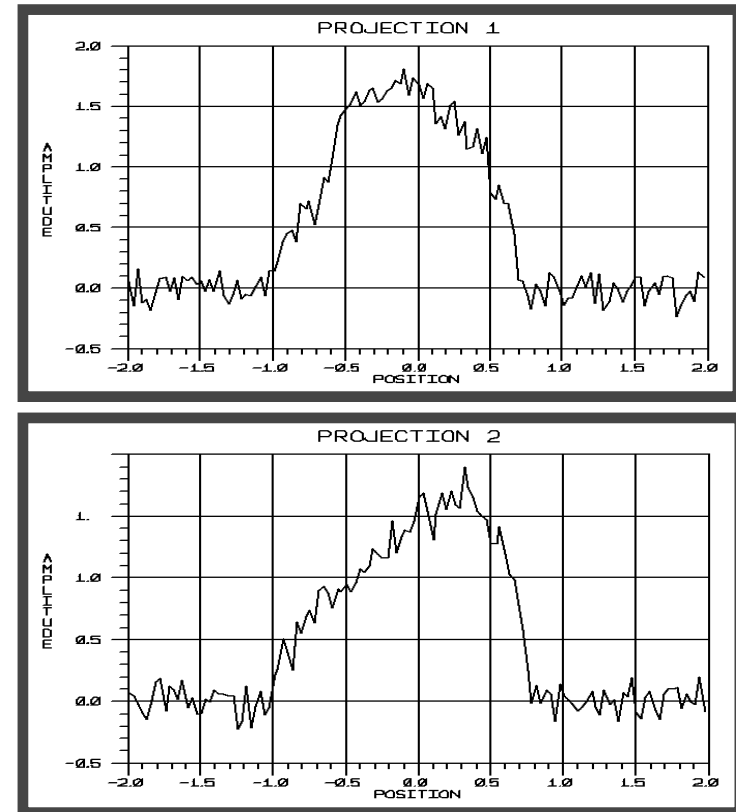
# Example - tomographic reconstruction

- Problem - reconstruct object from two projections
  - ▶ 2 orthogonal, parallel projections (128 samples in each view)
  - ▶ Gaussian noise;  
rms-dev 5% of proj. max

Original object



Two orthogonal projections  
with 5% rms noise



# Prior information used in reconstruction

---

- Assumptions about object
  - ▶ object density is uniform
  - ▶ abrupt change in density at edge
  - ▶ boundary is relatively smooth
- Object model
  - ▶ object boundary - deformable geometric model
    - relatively smooth
  - ▶ interior has uniform density (known)
  - ▶ exterior density is zero
  - ▶ only variables are those describing boundary

# Likelihood

---

- Probability of data  $\mathbf{d}$ , given model and parameters  $\mathbf{a}$
- For measurements degraded by independent Gaussian-distributed noise, minus-log-likelihood is

$$-\log[p(\mathbf{d}|\mathbf{a})] = \varphi(\mathbf{a}) = \frac{1}{2} \chi^2 = \frac{1}{2} \sum \frac{(d_i - d_i^*)^2}{\sigma^2}$$

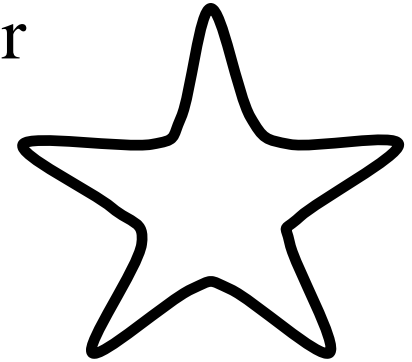
- ▶ where  $d_i$  is the  $i$ th measurement,  
 $d_i^*$  its predicted value (for specific  $\mathbf{a}$ ),  
 $\sigma$  is rms noise in measurements



# Deformable geometric models

---

- Natural to describe objects in terms of their boundaries
- In data analysis aim is to balance
  - ▶ internal energy  $\varepsilon$ : measure of deformation
  - ▶ external energy, e.g.  $\chi^2$ : measure of mismatch to data
- Constrain smoothness based on curvature  $\kappa$ 
  - ▶ deformation energy, e.g.,  $\varepsilon \sim \int \kappa^2 ds$ , for curve
  - ▶ controls number of degrees of freedom of curve
- Analogy to elastic materials - rods, sheets

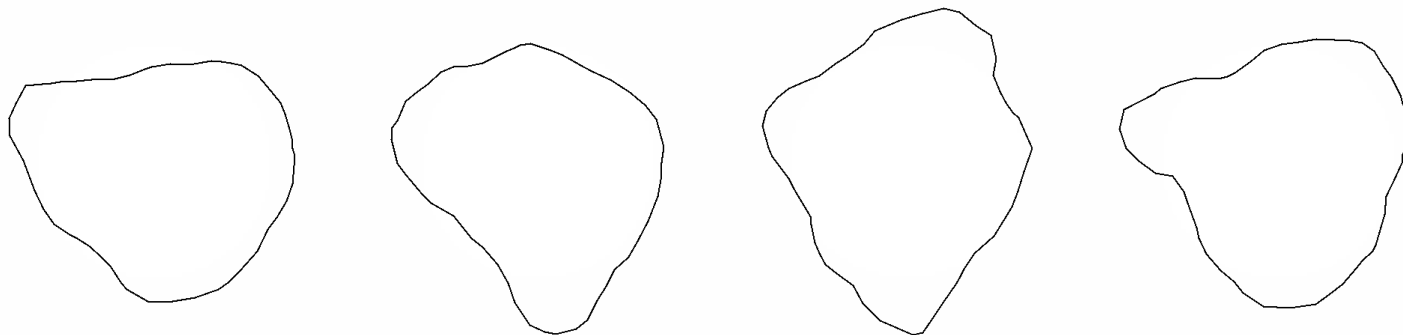


# Probabilistic interpretation of prior for deformable model

---

- Probability of shape:  $\sim \exp\left[-\frac{\alpha S}{(2\pi)^2} \oint \kappa^2 ds\right]$
- Sample prior pdf using MCMC
  - ▶ shows variety of shapes deemed admissible before experiment, capturing our uncertainty about shape
  - ▶ decide on  $\alpha = 5$  on basis of appearance of shapes

Plausible shapes drawn from prior for  $\alpha = 5$



# Tomographic reconstruction from two views

---

- Data consist of two orthogonal views
  - ▶ parallel projections, each containing 128 samples
  - ▶ Gaussian noise; rms-dev 5% of proj. max
- Object model
  - ▶ boundary is 50-sided polygon
  - ▶ smoothness achieved by prior on curvature  $\kappa$
  - ▶ uniform (known) density inside boundary
- $\varphi = -\log \text{posterior} = \frac{1}{2}\chi^2 + \frac{\alpha S}{(2\pi)^2} \oint \kappa^2 ds$  ,
  - ▶ where  $S$  is total perimeter,
  - ▶  $\chi^2$  is sum of squares of residuals divided by noise variance

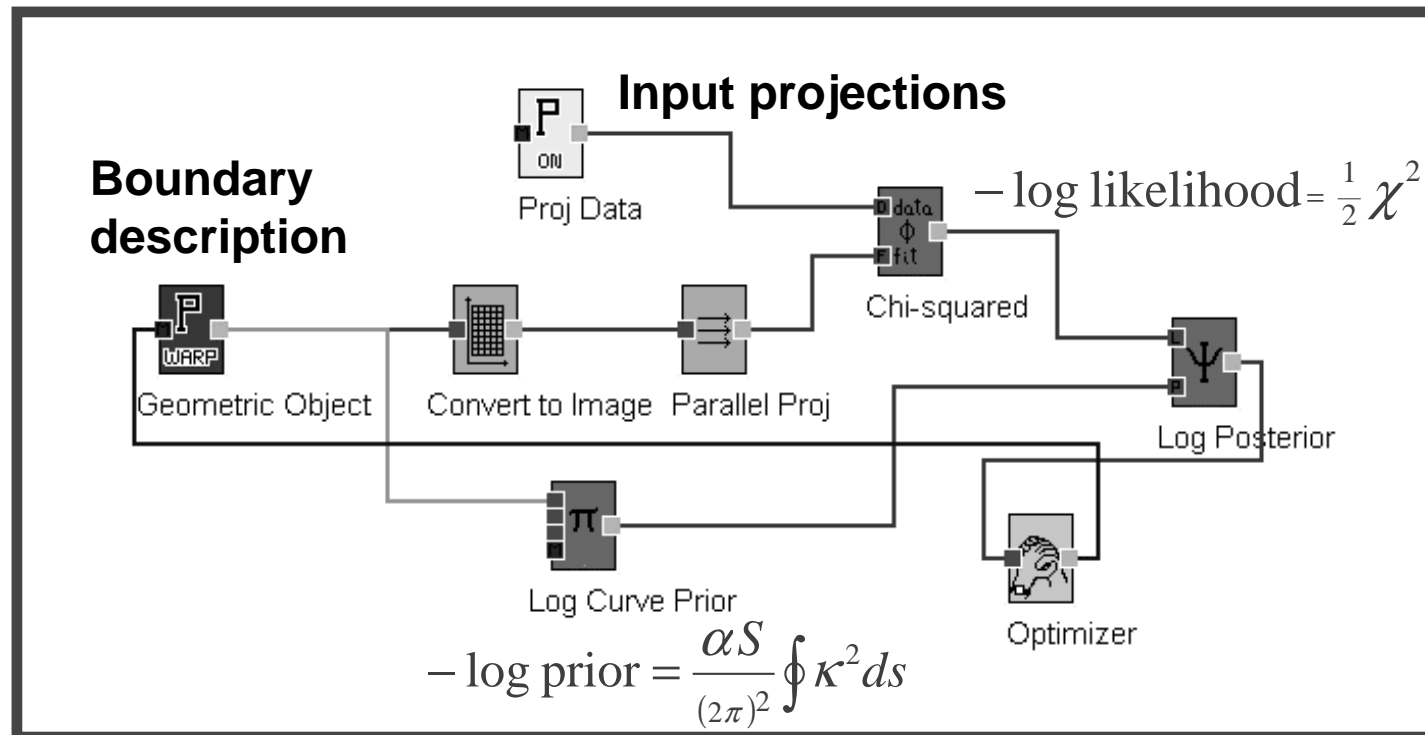
# The Bayes Inference Engine

---

- Flexible modeling tool developed at LANL
  - ▶ object described as composite geometric and density model
  - ▶ measurement process (principally radiography)
- User interface via graphically-programmed data-flow diagram
- Full interactivity with every aspect of model
- Provides
  - ▶ MAP estimate by optimization (gradient by adj. diff.)
  - ▶ samples of posterior by MCMC
  - ▶ uncertainty estimates

# The Bayes Inference Engine

- BIE data-flow diagram to find MAP solution

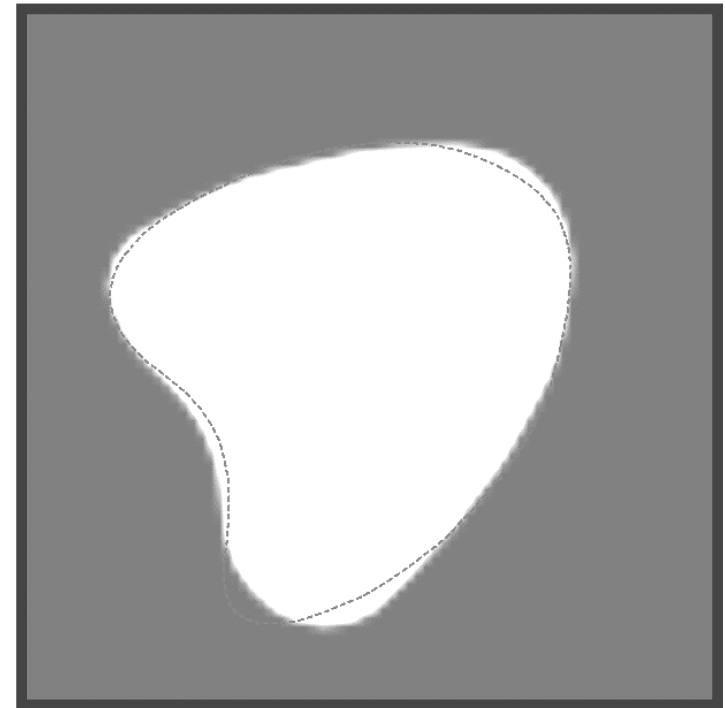


- Optimizer uses gradients that are efficiently calculated by adjoint differentiation in code technique(ADICT)

# MAP reconstruction

---

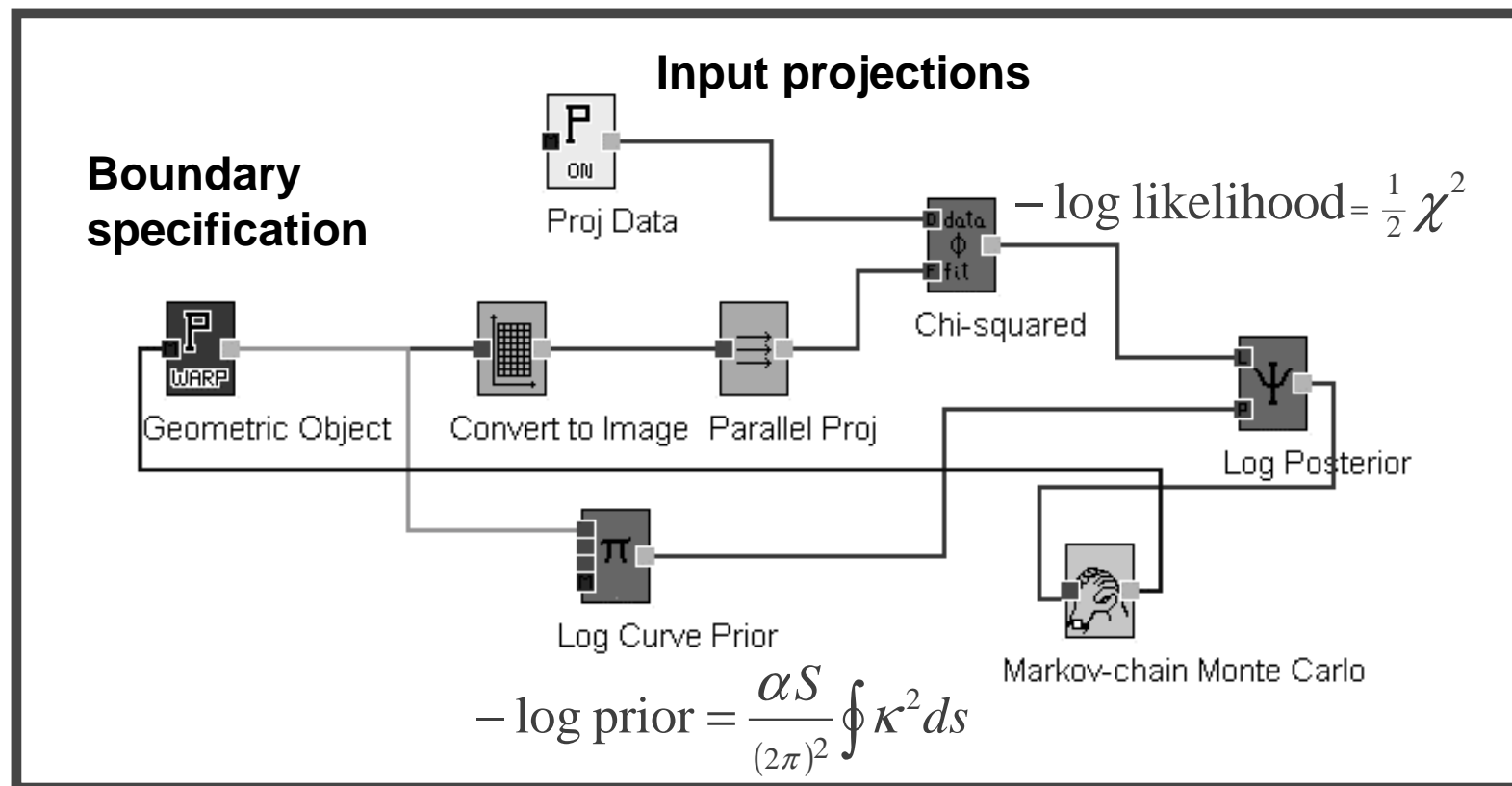
- Determine boundary that maximizes posterior probability
- Not perfect, but very good for only two projections
- Question: How do we quantify uncertainty in reconstruction?



**Reconstructed boundary  
(gray-scale) compared with  
original object (red line)**

# The Bayes Inference Engine

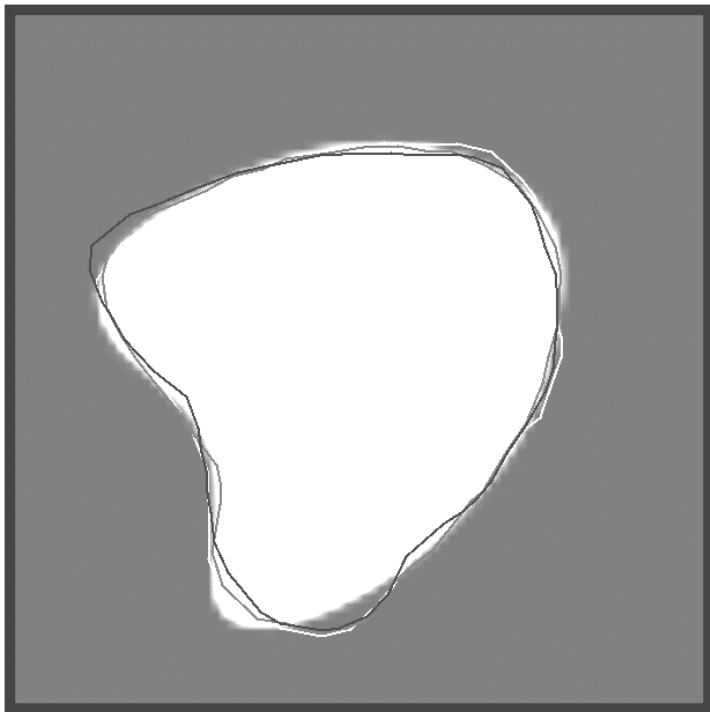
- BIE data-flow diagram to produce MCMC sequence



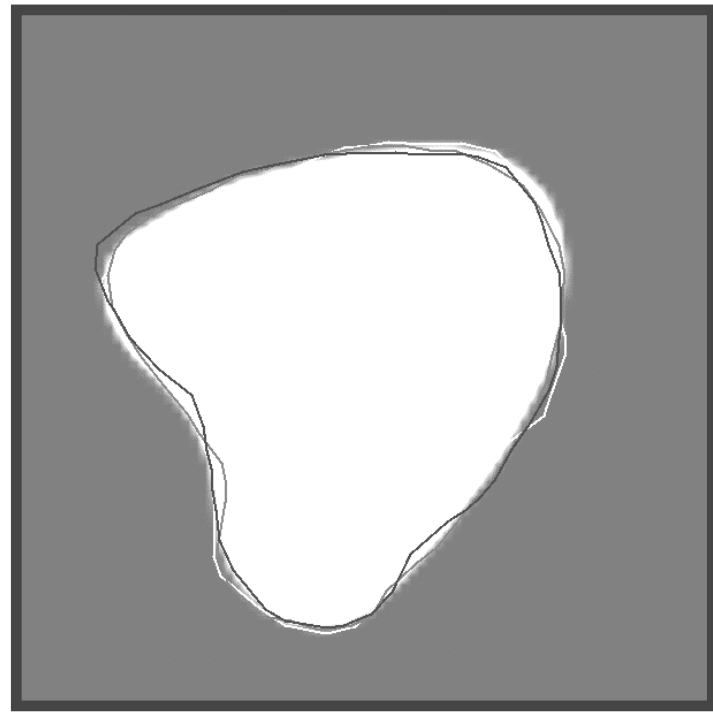
# Uncertainties in two-view reconstruction

---

- From MCMC samples from posterior with 150,000 steps, display three selected boundaries
  - these represent alternative plausible solutions



compared to original object



compared to MAP estimated object



# Visualization of uncertainty

---

- Problem inherently difficult for numerous parameters
  - ▶ wish to see correlations among uncertainties in parameters
- View MCMC sequence as video loop
  - ▶ advantage is one directly observes model in normal way
- View several plausible realizations from MCMC sequence
- Marginalized uncertainties (one parameter at a time)
  - ▶ rms uncertainty (or variance) for each parameter
  - ▶ credible intervals

# Posterior mean of gray-scale image

---

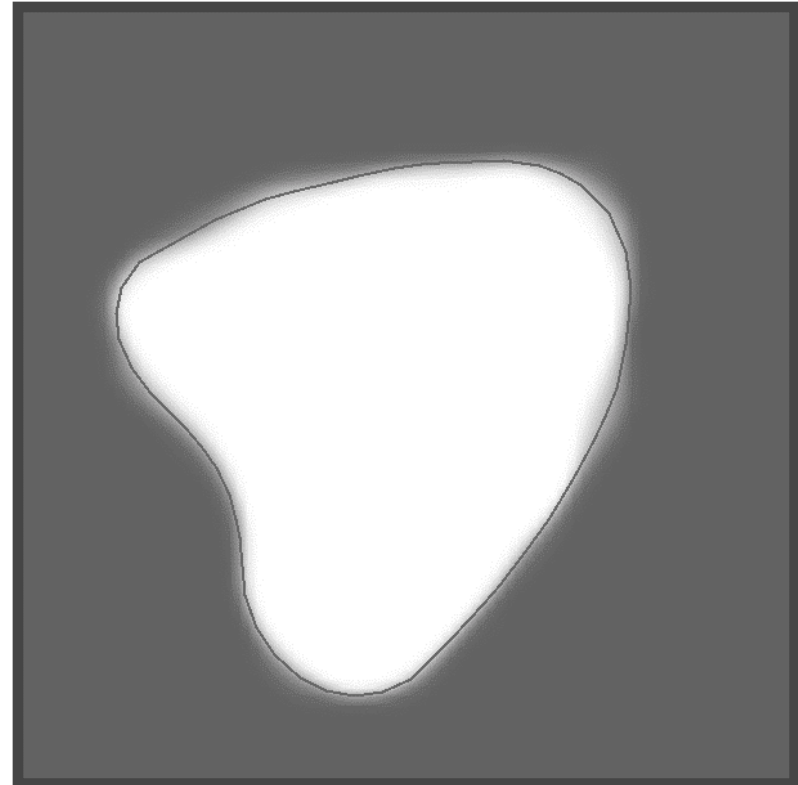
- Average gray-scale images over MCMC samples from posterior
- Value of pixel is probability it lies inside object boundary
- Amount of blur in edge is related to magnitude of uncertainty in edge localization
- Observe that posterior median nearly same as MAP boundary
  - ▶ implies posterior probability distribution symmetric about MAP parameter set

# Posterior mean of gray-scale image

---

- Pixels in posterior mean image with value 0.5 represent posterior median boundary position
  - ▶ similar to MAP boundary for two-view problem

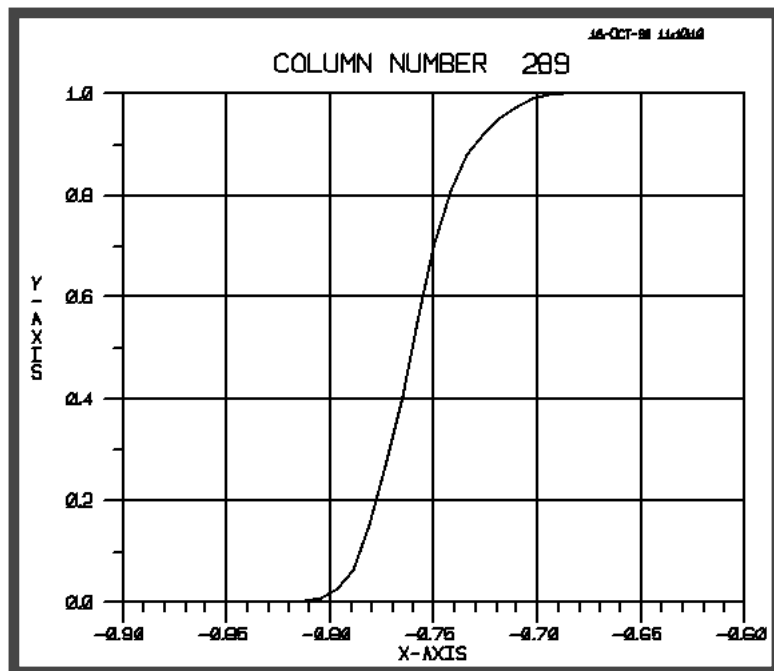
Posterior mean image  
compared to  
MAP boundary (red line)



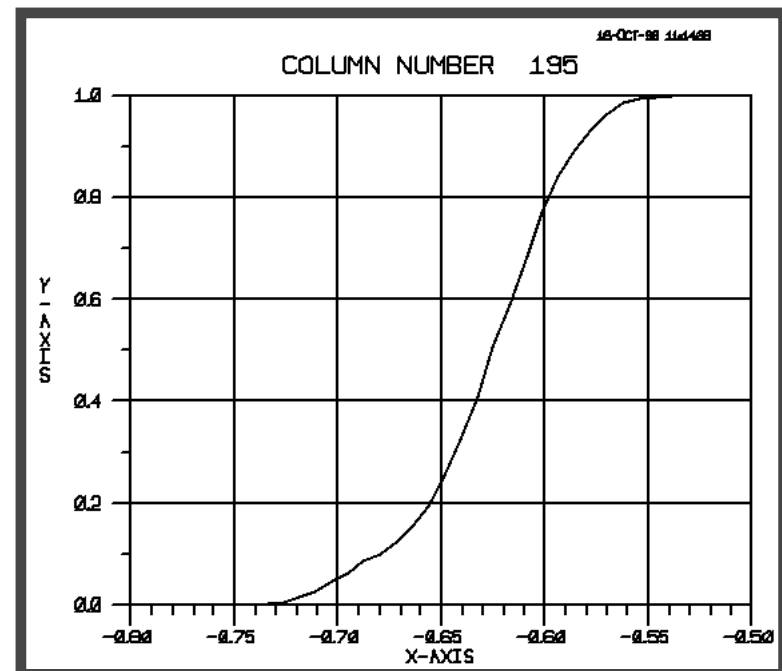
# Uncertainty in edge localization

- Steepness of edge profile of posterior mean image indicates uncertainty in edge localization
  - uncertainty is nonstationary; varies with position

At top of reconstruction (tangent point)



Top, left of center, less well determined



# Credible interval

---

- Bayesian "confidence interval"
  - ▶ probability that actual parameter lies within interval
  - ▶ different from standard definition of confidence interval, which is based on (hypothetical) repeated experiments
- For MCMC posterior mean image, determines credible interval for boundary position
  - ▶ 95% credible interval is region of posterior mean image in which pixel values lie between 0.025 and 0.975.

# Credible interval

---

- 95% credible interval of boundary localization for two-view reconstruction compared with original object boundary (red line)
  - ▶ narrower at tangent points
  - ▶ 92% of original boundary lies inside95% credible interval
- Marginalized measure of uncertainty - ignores correlations among different positions



# Important issues

---

- Bayesian vs. frequentist approach to uncertainty assessment
  - ▶ MCMC sampling of posterior
    - single data set, single object
  - ▶ Monte Carlo simulation of repeated experiments to determine characteristics of the estimator used
    - variety of data sets (variety of objects)
- Advantages of Bayesian approach
  - ▶ applies to the specific data set supplied
  - ▶ exposes null space; multiple solutions that yield exactly same measurements

# Stiffness of posterior

---

- Gaussian approximation for posterior:
  - $\log p(\mathbf{a}|\mathbf{d}) = \varphi = \varphi_0 + (\mathbf{a} - \mathbf{a}_0)^T \mathbf{K} (\mathbf{a} - \mathbf{a}_0)$ 
    - ▶ where  $\mathbf{a}$  is parameter vector
    - $\mathbf{K}$  is the curvature or second derivative matrix of  $\varphi$  (aka Hessian) and
    - $\mathbf{a}_0$  is the position of the minimum in  $\varphi$  (MAP estimate)
- Curvature matrix  $\mathbf{K}$  is measure of stiffness of solution
- Covariance matrix is inverse of  $\mathbf{K}$ :  $\mathbf{C} = \mathbf{K}^{-1}$



# Determining stiffness of posterior

---

- First estimate  $\mathbf{a}_0$  by minimizing  $\varphi$  (MAP solution)
- Apply force to model (a vector in parameter space)
- Effect of force is to add potential to  $\varphi$ :

$$\varphi = \mathbf{f}^T (\mathbf{a} - \mathbf{a}_0) + \varphi_0 + (\mathbf{a} - \mathbf{a}_0)^T \mathbf{K} (\mathbf{a} - \mathbf{a}_0)$$

- Minimizing  $\varphi$  again; setting gradient of  $\varphi$  to zero

$$\mathbf{K} (\mathbf{a} - \mathbf{a}_0) = \mathbf{f}$$

$$\text{or} \quad \mathbf{a} - \mathbf{a}_0 = \mathbf{K}^{-1} \mathbf{f} = \mathbf{C} \mathbf{f}$$

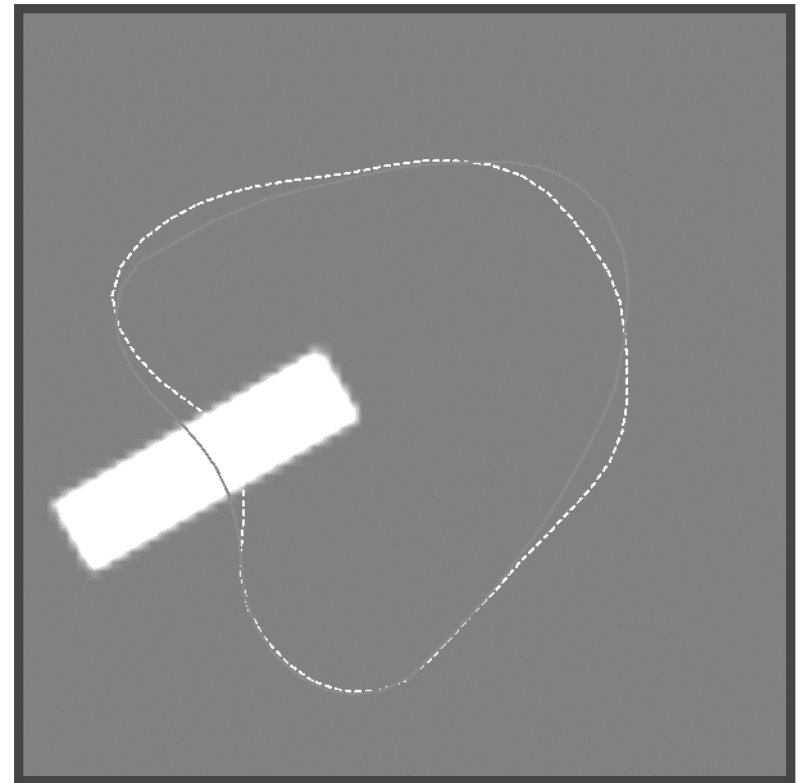
- Parameter displacement from MAP solution is proportional to covariance matrix times applied force
- We have called this the “hard truth” method, because truth is hard!

# Hard truth method

---

- Interpret  $\varphi = -\log$  probability as potential function; sum of
  - deformation energy
  - $\frac{1}{2}\chi^2$
- Stiffness of model proportional to curvature of  $\varphi$
- Row of **covariance matrix** is displacement obtained by applying a force to MAP model and reminimizing  $\varphi$

Applying force (white bar) to MAP boundary (red) moves it to new location (yellow-dashed)



# Summary

---

- MCMC technique - to sample arbitrary pdf
- Bayesian approach to model building
  - ▶ uncertainty assessment
  - ▶ MCMC sampling of posterior
    - covariance estimates
    - credible intervals
  - ▶ permits use of prior information
- Deformable geometric models
  - ▶ useful to capture notions of object shape
  - ▶ smoothness prior states preference for smooth boundary
- Example - tomographic reconstruction from two views

# Bibliography - MCMC

---

- ▶ “Posterior sampling with improved efficiency,” K. M. Hanson and G. S. Cunningham, *Proc. SPIE* **3338**, 371-382 (1998); includes introduction to MCMC
- ▶ *Markov Chain Monte Carlo in Practice*, W. R. Gilks et al., (Chapman and Hall, 1996); excellent general-purpose book
- ▶ “Efficient Metropolis jumping rules,” A. Gelman et al, in *Bayesian Statistics* **5**, J. M. Bernardo et al., (Oxford Univ., 1996); diffusion theory
- ▶ “Bayesian computation and stochastic systems,” J. Besag et al., *Stat. Sci.* **10**, 3-66 (1995); MCMC applied to image analysis
- ▶ *Bayesian Learning for Neural Networks*, R. M. Neal, (Springer, 1996); Hamiltonian hybrid MCMC
- ▶ “Bayesian multinodal evidence computation by adaptive tempered MCMC,” M-D Wu and W. J. Fitzgerald, in *Maximum Entropy and Bayesian Methods*, K. M. Hanson and R. N. Silver, eds., (Kluwer Academic, 1996); annealing
- ▶ “Inversion based on complex simulations,” K. M. Hanson, *Maximum Entropy and Bayesian Methods*, G. J. Erickson et al., eds., (Kluwer Academic, 1998); describes adjoint differentiation and its usefulness

# Bibliography - Bayesian analysis

---

- "Uncertainty assessment for reconstructions based on deformable models," K. M. Hanson et al., *Int. J. Imaging Syst. Technol.* **8**, pp. 506-512 (1997)
- "The hard truth," K. M. Hanson et al., in *Maximum Entropy and Bayesian Methods*, pp. 157-164 (Kluwer, 1996)
- "Operation of the Bayes Inference Engine," K. M. Hanson et al., in *Maximum Entropy and Bayesian Methods*, pp. 309-318 (Kluwer, 1999)
- "Inference from Rossi traces," K. M. Hanson et al., to be published in *Maximum Entropy and Bayesian Methods*, (Kluwer, 2001)

These and other papers available under <http://home.lanl.gov/kmh/>